Small-scale Forestry (2014) 13:117–125
DOI 10.1007/s11842-013-9244-4

RESEARCH PAPER

Forest Fire Risk Zone Modeling Using Logistic Regression and GIS: An Iranian Case Study

Frouzan Mohammadi · Mahtab Pir Bavaghar · Naghi Shabanian

Accepted: 28 May 2013 / Published online: 4 June 2013

Abstract Forest fires are an important environmental concern worldwide, affecting the soil, forests and human lives. During the process of burning, soil nutrients are depleted and the soil is subsequently more vulnerable to erosion. Nowadays it is necessary to identify the factors influencing the occurrence of fire and fire hazard areas, in order to minimize the frequency of fire and avert damage. Logistic regression was used to study the forest fire risk and identify the most influential factors in the occurrence of forest fires. Climatic variables (temperature and annual precipitation), human factors (distance from streams and farmland) and physiography (land slope and elevation) were considered and their correlation with the occurrence of fires investigated. Results of model validation and sensitivity of various areas to fire were examined with the ROC coefficient and Hosmer–Lemeshow test. The estimated coefficients for the independent variables indicated that the probability of occurrence of fire is negatively related to land slope, site elevation and distance from farmlands, but is positively related to amount of annual precipitation.

Keywords Risk map · Physiography · Climate · Validation · ROC curve

F. Mohammadi · N. Shabanian
Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
e-mail: mohamadi.arshad89@yahoo.com

N. Shabanian
e-mail: n.Shabanian@uok.ac.ir

M. P. Bavaghar (✉)
Center for Research and Development of Northern Zagros Forests, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran
e-mail: mahtab_bavaghar@yahoo.com; m.bavaghar@uok.ac.ir

Introduction

Forests are a major natural resource that have an important role in retaining environmental balance. The vitality of a forest is a prevailing indicator of the ecological conditions in the area. One of the reasons for the destruction of forests throughout the world is frequent incidence of fire (Kandya et al. 1998). Frequent occurrence of forest fires is a major threat for natural resources and even human lives. Therefore predicting the risk of forest fire is critical for maintaining forest resources. Forest fires are a potential risk with physical, biological, ecological and environmental consequences (Johnson and Gutsell 1994; Jaiswal et al. 2002). Fire threatens the standing vegetation, wild animals, small trees and forest regeneration. Ground fire destroys organic material, which is needed to retain humus in the soil. Annual fires and consequent decline of the growth of the grasses, herbs and shrubs, cause soil erosion (Kandya et al. 1998). Due to these negative impacts, fire management is required to protect forests. Lack of trusted records about occurrence of fire and its spatial distribution is a critical topic for fire management (Chuvieco and Congalton 1989; Chuvieco and Salas 1996; Chuvieco 1999). A forest fire can become a major ecological tragedy regardless of what caused the fire (natural factors or human activities). Forest fire risk zone mapping is necessary to decrease the frequency of fire and avert harm (Dong et al. 2005). Several environmental factors, including fuel load (vegetation cover), climate condition and physiography (elevation and slope), have major impacts over the creation, propagation and intensity of forest fires (Brown and Davis 1973).

A major problem for forest management is the lack of information on affected areas, including recent fire frequencies (Geldenhuys 1996). At all spatial scales, satellite remote sensing has opened up opportunities to monitor and evaluate forests and other ecosystems. Remote sensing has been used in this study for monitoring and detection of forest fires.

Many fire risk models have been developed based on environmental factors that influence wildfire. A major part of forest fire management planning is identifying areas that have a high probability of forest fire occurrence. GIS models are useful to assist managers for mapping and analyzing the variables that contribute to the occurrence of fire across large, unique geographic units. The objective of this study was to generate a forest fire risk zone map, and to utilize GIS coupled with spatial logistic regression analysis to define the relationship between physiographic and climate characteristics and human activities related to forest fire patterns in western Iran.

Research Method

The study area is Sarvabad forests in Kurdistan province of western Iran. This area is located between latitude 46°15′N to 46°35′N and longitude 35°15′E to 35°30′E (Fig. 1), and has an elevation ranges from 900 to 2,400 masl. The total area of the study region is 11,336, 6,200 ha of which is covered by natural forest. Wildfire had occurred repeatedly in this area, and field data are available for 2006–2011. The fire map was generated according to this existing information, and a field reconnaissance survey.

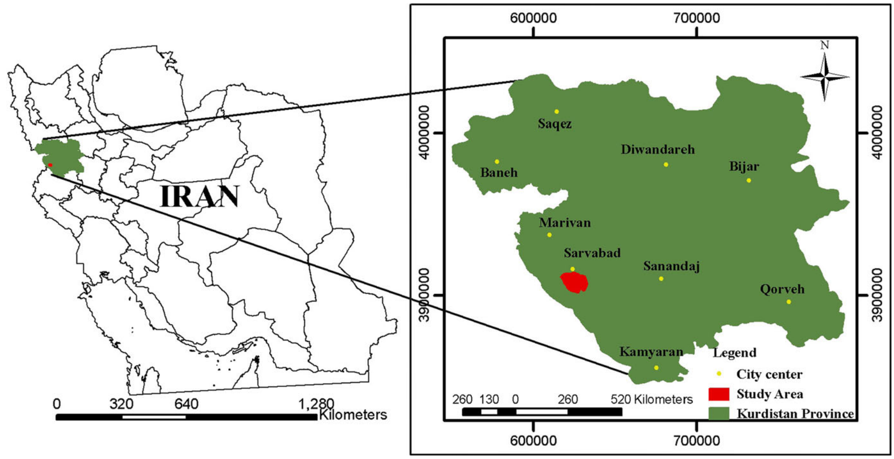

Fig. 1 The study area of Sarvabad, Kurdistan, Iran

A vegetation density map at the scale of 1:250,000 was available, and was edited and updated using SPOT5 data acquired in 2005. Forest fire records relating to 2006–2011 were obtained from the general office of natural resources of Kurdistan. A topographic map at the scale of 1:25,000 was used to generate a digital elevation model (DEM). Meteorological data of 2006–2011 were also obtained to generate the maps of climate factors.

Generation of Thematic Maps

It was hypothesized that forest fire risk is related to canopy density, physiographic features (elevation and slope), climate variables (temperature and amount of precipitation) and distance to streams and distance to farmlands. The vegetation density layer was obtained using a vegetation density map and was updated using inventory and SPOT5 satellite image interpretation.

Physiographic factors were chosen because variations in them could result in dramatic changes in fire behaviour. The DEM was produced using a topographic map at the scale of 1:25,000. The slope and elevation maps were extracted from digital elevation model.

Distance of locations in meters to streams and farmlands obtained from the digital maps and site surveying. These maps were converted from vector to raster format with 20 m grid cells.

Records of annual precipitation and temperature for the period 2006–2011 were obtained from a meteorological bureau. A temperature layer was derived based on an estimated two-variable regression models between average monthly temperature in August and elevation. August was chosen because this is the warmest month in this area and has the most frequent fires. Spatial interpolation of annual precipitation was carried out using existing datasets to generate the rainfall distribution layer for

the study area. In this process, one of the most common interpolation methods—called 'inverse weights distance'—was employed. This method assumes that the attribute value of an unsampled point is the weighted average of known values within the neighborhood, and the interpolating surface is influenced most by the nearest points.

The geographic information system in ArcGIS software has been used to transfer the variables of vegetation density, physiography, human activities and climate into a database for determination of forest fire risk. These thematic maps have been classified according to objective, accuracy and scale.

Modeling the Spatial Distribution of Forest Fire

Logistic regression has been used for determining weights of variables as well as to investigate relationships between occurrence of forest fire and explanatory variables. In the investigation of forest fire risk assessment, fire presence (hotspots) is the dependent variable, while the environmental and human factors are the independent variables.

One hundred points were selected at random in forest areas where fire had occurred between 2006 and 2011, and 100 points were selected in forest areas where fire had not occurred over the same period. To decrease the spatial autocorrelation, the points should be separated by distance of at least 1,000 m (Koenig 1999). However, the burned areas were composed of small polygons, and therefore at least 100 m separation distance between samples was judged to be adequate.

The spatial characteristics of each of the sample points were obtained by extracting data from the elevation, slope, temperature and annual precipitation layers as well as maps of the distance to streams and farmland. These data were imported into SPSS. Because the explanatory variables are measured at different scales, they do not contribute equally to the analysis, making it difficult to assess relative importance. Transforming the data to comparable scales can prevent this problem. Hence the explanatory variables were standardized by dividing values by their root-mean-square (following Etter et al. 2006). A binary logistic regression model was then used to determine which of the factors determined the probability of occurrence of forest fire, and a prediction model for it. Eighty percent of data points were used for modeling and 20 % for model validation. A logistic model was estimated based on one binary dependent variable (fire presence = zero, non-presence fire = 1) and six independent variables (elevation, slope, distance to streams, distance of farmlands, temperature and annual precipitation). The hypothesized population logistic model has the form:

$$P = E(Y) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i)}$$

where P is the probability that a fire occurs, β_0 is the intercept and the β_i are the slope parameters associated with the independent (X_i) variables.

The predictive efficiency of the model was assessed by using the model to predict the probability of forest fire and by comparing the predictions with the actual distribution of fires in the model validation datasets. Discrimination is a relevant measure of predictive performance when predictions are needed for arranging areas according to their relative vulnerability (Hanley and McNeil 1982; Schneider and Pontius 2001), because the discrimination ability of a model provides information on the degree to which higher predicted probabilities are associated with the presence of forest fire. The area under the receiver operating characteristic (ROC) curve was employed to assess the discrimination ability of the model. The ROC curve tests the model ability's to predict accurately a binary response, given the values of the predictors in the estimated model. ROC curve values range from 0.5 to 1.0, with larger values indicative of better fit (Wilson et al. 2005). The coefficient of determination (R^2) measure is only appropriate to linear regression, with its continuous dependent variables. For logistic regression, a number of statisticians have developed so-called 'Pseudo R^2' measures which take a different conceptual approach but aim to mimic R^2 for logistic regression models. Due to the binary nature of dependent variable (fire presence/absence), the pseudo R^2 measure will tend to be lower than traditional ordinary least squares R^2 measure (Bio et al. 1998), and cannot be a good indicator for conventional logistic regression analysis (Wilson et al. 2005). The Wald and Chi square tests are used to examine the statistical significance of the individual regression coefficients (β_i). Negative β_i coefficients indicate negative correlation between dependent and independent variables. The Hosmer–Lemeshow test was employed in this study for testing goodness-of-fit. If the Hosmer–Lemeshow test statistic is >0.05, the null hypothesis that the model is fit cannot be rejected.

Several models derived using logistic regression analysis. The best model was selected based on results of validation and accuracy assessment of the models. The probability of forest fire occurrence is mapped using the selected model.

Results

The locations of forest fires were determined with field inventory and forest fire records. The area of forest fires explained approximately 3.3 % of the forest area that was burnt during the 5-year study period (2006–2011). Therefore, the mean annual fires rate was 0.67 %.

The primary information about the occurrence of forest fires related to fire-influencing factors is presented in Table 1. The dependent variable is binary (fire presence/absence) and the predictors variables (independent variables) are the six factors discussed above (elevation, slope, distance from streams to farmlands, temperature and annual precipitation).

The probability of forest fire was significantly and negatively related to elevation (Wald statistic = 5.86, $p < 0.05$), slope (Wald statistic = 3.70, $p < 0.05$) and distance from farmlands (Wald statistic = 12.63, $p < 0.01$). Occurrence of forest fire was positively related to annual precipitation (Wald statistic = 4.22, $p < 0.05$).

Table 1 The occurrence of forest fire related to influencing factors

Variable	Classes	Fire occurrence (%)
Elevation (m)	900–1,300	39
	1,300–1,700	57
	1,700–2,100	4
	2,100–2,500	0
Slope (%)	0–25	46
	25–50	37
	50–75	16
	>75	1
Distance from farmland (m)	0–300	31
	300–600	28
	600–900	21
	900–1,200	17
	>1,200	3
Distance from stream (m)	0–100	8
	100–200	14
	200–300	16
	300–400	15
	>400	46
Temperature (deg C)	16–25	0
	25–30	48.5
	30–42	51.5
Annual precipitation (mm)	730–770	0
	770–810	27.6
	810–850	72.4

The estimated forest fire risk model explained 70 % of the variation in the recorded observations, had an ROC value of 0.794 and was not affected by spatial autocorrelation. The logistic regression goodness of fit measured by the Nagelkerke R^2 statistic is 0.304. A Hosmer–Lemeshow statistic of 0.20 was obtained, which is not statistically significant, so the model is fit. The coefficients and values of the logistic regression model as well as validation results are presented in Table 2.

The forest fire probability map based on the binary logistic regression model is presented in Fig. 2. The best-fit model of forest fire risk was selected to predict the probability of forest fire risk. The probability of forest fire risk ranged from 0 to 0.99. The areas with low elevation, low slope, short distance from farmland to higher precipitation have higher values in fire probability map and therefore are more prone to forest fire. Rocky areas with high elevation take a value of zero. The probability threshold of 0.5 explained the best agreement between areas predicted to develop fires and areas where forest fires have actually occurred. Therefore, areas with a predicted probability of fire occurrence >0.5 were seen to be highly prone to fire occurrence.

Table 2 Explanatory variables and significance levels for the fire risk model

Variable[a]	Test statistics				
	B coefficient	Standard error	Wald	df	*p* level
Elevation	−0.094	0.039	5.858	1	<0.05
Slope	−0.032	0.016	3.967	1	<0.05
Distance to farmland	−0.048	0.013	12.626	1	<0.01
Annual precipitation	0.092	0.045	4.221	1	<0.05
Constant	−21.124	13.207	2.65	1	<0.1

Chi square value = 42.385

Nagelkerke $R^2 = 0.304$

ROC = 0.794, SE = 0.031 p level <0.001

Hosmer–Lemeshow test Chi square = 11.03, p value = 0.20

Correct classification (correct estimation of predicted values) = 70 %

Model used: fire occurrence = −21.124 to 0.09 elevation −0.032 slope −0.048 distance of farmlands +0.092 precipitation

[a] Standardized values of variables

Discussion and Conclusion

A reliable way to identify the susceptibility of particular areas to fire and to create a fire risk model can be the combination of GIS and logistic regression analysis. Such a model is reported in this paper, in which the forest fire risk probability map is formed according to elevation, slope, distance to farmlands and annual rainfall variables.

The fire risk in this study reflects both the likelihood of occurrence and the risk of spreading of the fire. The slope factor influences the risk of spreading. Fire scars in relation to slope were observed in the low slope classes rather than higher slope classes. Low slopes are favourable for agricultural use, so are managed carefully by humans for food production. Sometimes, farmers ignite fires in forests to increase their cropping area (Jaiswal et al. 2002; Dong et al. 2005). Elevation affects the length of fire season and land cover (Pyne et al. 1996). Forest fires records have shown that 96 % of fires occurred at elevations below 1,700 m (Table 1). The probability of forest fire ignition is higher at lower elevation due to higher temperature and lower rainfall. Another reason is that there in the Sarvabad forests is the relatively high human population at elevations below 1,700 m. These results are consistent with the finding of other research, including Setiawan et al. (2004) and Dong et al. (2005).

Human activities have spatial relationships with position of fires caused by humans. Human factors are most closely related to forest fire incidence at locations closest to farmland. Farms are mainly located in sunny areas at lower elevation, so forest fires often occur in these areas. Farmland is located near forests in Sarvabad, therefore human and cars movement is high in these areas causing higher probability for forest fire. These results are consistent with the finding of other research, including Kalabokidis et al. (2002) and Dong et al. (2005). There was only

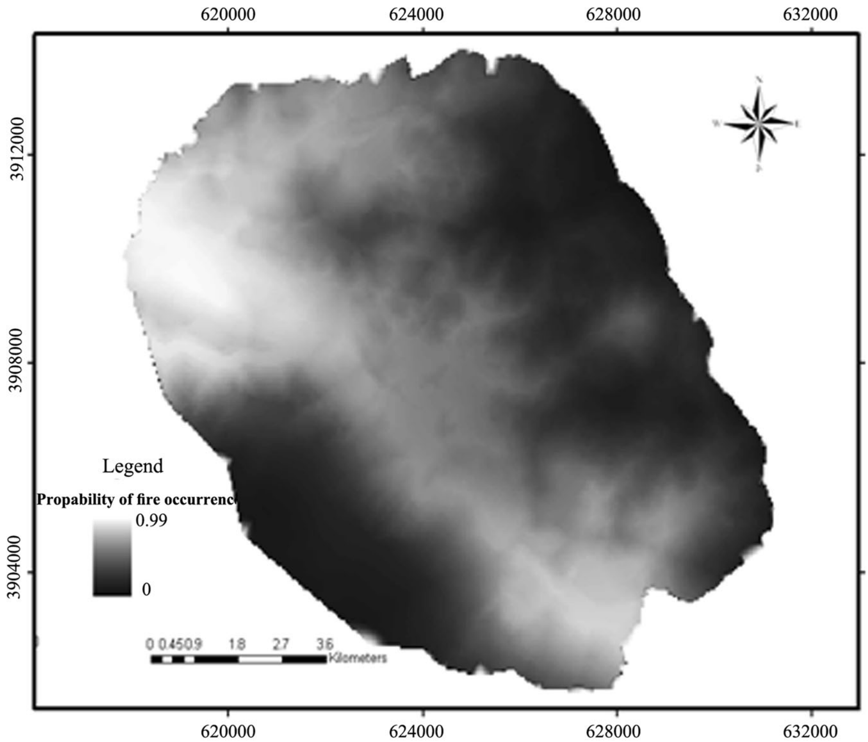

Fig. 2 Probability of forest fire occurrence in the study area, according to the logistic regression model. *Note* The *lighter areas* have a higher probability of forest fire occurrence, are located in low elevation, and are relatively flat areas. These areas, also have lower distance to farmlands and higher precipitation

weak correlation between distance and streams and forest fire in the study area. Most of the land located near streams was allocated to vegetable gardens, and was therefore protected by the landowners. Temperature is a function of elevation, so when elevation entered into the model, the temperature variable was removed to avoid multicollinearity.

According to the model, there is a positive relationship between probability of fire and annual precipitation. It was found that 72 % of fires have occurred within the region with more than 800 mm of average annual rain. Higher rainfall leads to further growth of grasses on the forest floor. In such circumstances, with drying of floor cover grasses in summer, the incidence of fire occurrence increases in wetter sites.

The model validation confirmed that the independent variables have adequate power to discriminate between burned and non-burned areas as well as to predict forest fire risk. Furthermore, the quality and accuracy of the maps of the explanatory variables included in the model appear adequate.

The model quality could be improved if further variables that may affect the forest fire are imported into the logistic regression analysis. The relationships between variables may change over time, so periodic updating the model is desirable.

This research demonstrates that logistic regression modeling and geographic information system are suitable for determining the forest fire risk zone, and therefore management of it. The analysis has revealed that the elevation, slope, annual precipitation and distance to farms have high significant correlation with fires. The logistic regression method combined with GIS and inventory data is useful for fire risk mapping, because each factor influencing forest fire risk is analyzed in this method. This modelling and mapping provides valuable information about areas most likely to be affected by fire. Forest fire risk zone mapping is a useful tool in forest fire prevention and management in order to minimize wildfire risk and damage, allowing forest fire managers to identify high fire risk locations easily and manage these areas effectively.

References

Bio AMF, Alkemende R, Barendregt A (1998) Determining alternative models for vegetation response analysis: a non-parametric approach. J Veg Sci 9:5–16

Brown A, Davis KP (1973) Forest fire: control and use, 2nd edn. McGraw-Hill, New York

Chuvieco E (ed) (1999) Remote sensing of large wildfires in the European mediterranean basin. Springer, Berlin

Chuvieco E, Congalton RG (1989) Application of remote sensing and geographic information systems to forest fire hazard mapping. Remote Sens Environ 29(2):147–159

Chuvieco E, Salas J (1996) Mapping the spatial distribution of forest fire danger using GIS. Int J Geogr Inf Sci 10(3):333–345

Dong X, Li-min D, Gue-fan Sh, Lei T, Hui W (2005) Forest fire risk zone mapping from satellite images and GIS for Baihe forestry Bureau, Jilin, China. J For Res 16(3):169–174

Etter A, McAlpine C, Wilson K, Phinn S, Possingham H (2006) Regional patterns of agricultural land use and deforestation in Colombia. Agric Ecosyst Environ 114:369–386

Geldenhuys CJ (1996) Options for sustainable harvesting of timber products from wood lands: examples from southern Africa. In: Proceedings of conference sustainable management of indigenous forests in the dry tropics, Forestry Commission, Harare, Kadoma, Zimbabwe, pp 124–142

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36

Jaiswal R, Mukerjee S, Raju K, Saxana R (2002) Forest fire risk zone mapping from satellite imaging and GIS. Int J Appl Earth Obs Geoinfo 4:1–10

Johnson EA, Gutsell SL (1994) Fire frequency models, methods and interpretations. Adv Ecol Res 25:239–287

Kalabokidis K, Konstatinidis P, Vasilkos Ch (2002) GIS Analysis of physical and human impact on wild fire patterns. In: viegas (ed) Forest fire research and wild land fire safety. Millpress, Rotterdam

Kandya AK, Kimothi MM, Jadhav RN, Agarwal JP (1998) Application of GIS in identification of fire prone areas—a feasibility study in parts of Junagarh (Gujrat, India). Indian For 124(7):531–535

Koenig WD (1999) Spatial autocorrelation of ecological phenomena. Trends Ecol Evol 14:22–26

Pyne SJ, Andrews PL, Laven RD (1996) Introduction to wildland fire, 2nd edn. Wiley, Toronto

Schneider LC, Pontius RG (2001) Modelling land-use changes in the Ipswich watershed, Massachusetts. USA Agric Ecosyst Environ 85:83–94

Setiawan I, Mahmud A, Masnsor S, Shariff A, Nuruddin A (2004) GIS-grid-based and multi-criteria analysis for identifying and mapping peat swamp forest fire hazard in Pahang, Malaysia. Disas Prev Manage J 13:379–386

Wilson k, Newton A, Echeverria C, Weston Ch, Burgman M (2005) A vulnerability analysis of the temperate forests of south central Chile. Biol Conserv 122(1):9–21